

MIT-CTP-2449
 TUTP-95-2
 hep-th/9507035
 June 1995

Quantum gravity effects at a black hole horizon*

Gilad Lifschytz

*Center for Theoretical Physics, Laboratory for Nuclear Science
 and Department of Physics,
 Massachusetts Institute of Technology, Cambridge MA 02139, USA.
 e-mail: gil1@mitlns.mit.edu*

and

Miguel Ortiz

*Institute of Cosmology, Department of Physics and Astronomy,
 Tufts University, Medford, MA 02155, USA.
 e-mail: mortiz@cosmos2.phy.tufts.edu*

Abstract: Quantum fluctuations in the background geometry of a black hole are shown to affect the propagation of matter states falling into the black hole in a foliation that corresponds to observations purely outside the horizon. A state that starts as a Minkowski vacuum at past null infinity gets entangled with the gravity sector, so that close to the horizon it can be represented by a statistical ensemble of orthogonal states. We construct an operator connecting the different states and comment on the possible physical meaning of the above construction. The induced energy-momentum tensor of these states is computed in the neighbourhood of the horizon, and it is found that energy-momentum fluctuations become large in the region where the bulk of the Hawking radiation is produced. The background spacetime as seen by an outside observer may be drastically altered in this region, and an outside observer should see significant interactions between the infalling matter and the outgoing Hawking radiation. The boundary of the region of strong quantum gravitational effects is given by a time-like hypersurface of constant Schwarzschild radius r one Planck unit away from the horizon. This boundary hypersurface is an example of a stretched horizon.

* This work was supported in part by funds provided by the U.S. Department of Energy (D.O.E.) under cooperative agreement DE-FC02-94ER40818, and in part by the National Science Foundation.

1 Introduction

The standard derivation of the Hawking radiation of a black hole involves the propagation of free fields on an idealised collapsing black hole background [1]. This simple calculation indicates that black holes evaporate through approximately thermal radiation. Unless there are subtle correlations between the matter that forms the black hole and the outgoing radiation that are ignored in this calculation, there is a net loss of information after the evaporation is complete.

Correlations between the degrees of freedom that form the black hole and those of the Hawking radiation appear to be ruled out by simple arguments [2]. According to the equivalence principle, matter freely falling into the black hole should not be subject to any extraordinary interactions as it passes through the event horizon. Thus the information carried by the infalling matter is carried into the black hole interior, and becomes inaccessible to an outside observer. Unless the information can be simultaneously contained in parts of the quantum state inside and outside the black hole, it is lost behind the event horizon.

Ever since Hawking's original derivation, there have been attempts to show that quantum gravity should affect the evaporation process in such a way that information is in fact conserved [3]. The preceding arguments indicate that if quantum gravity is to play a role in resolving the black hole information paradox, its effects must become important at the event horizon. However, the outgoing Hawking state is in the Unruh or Kruskal vacuum close to the horizon, and quantum field theory calculations have clearly shown that there are no large curvature or back-reaction effects there [4]. The only exception comes at the very late stages of evaporation, when the black hole becomes Planck-sized, and the receding horizon approaches the strong curvature region within the black hole core. However, at this late stage, it is very difficult to arrange for quantum gravity effects to restore information within a finite time [5].

Over the last few years, a number of new suggestions have been made of how non-perturbative quantum gravitational effects might be responsible for new physics at the black hole horizon. The most ardent proponent of this viewpoint has been 't Hooft [6] who as early as 1985 suggested that interactions between infalling matter and the outgoing Hawking radiation should not be ignored, since they involve Planckian energy scattering processes which can only be computed in a quantum theory of gravity. More recently several authors have added pieces to the puzzle, and a coherent picture of how quantum gravity could be important close to a black hole horizon has begun to emerge [6-13].

An important conceptual step was taken by 't Hooft and Susskind in introducing the notion of black hole complementarity. Susskind *et. al.* [7] point out that large quantum gravitational effects close to the horizon are only consistent if coordinate transformations involving boosts on the order of the Planck scale, transform the matter states in an unconventional way. It is conjectured that in this way, freely-falling observers could experience no large effects near the horizon, while outside observers at rest with respect to the horizon perceive it as a region of strong interaction where

semiclassical physics breaks down and infalling states interact significantly with the outgoing radiation.

A series of calculational frameworks have emerged which allow the quantum gravitational effects predicted by 't Hooft to be computed and which indicate the possibility of corrections to the standard Hawking derivation. It was first noted by 't Hooft [6] that the effect of planckian scattering between infalling and outgoing matter was to cause a shift in Kruskal coordinates $x^- \rightarrow a^-(x^- + \Delta^-)$ in the outgoing Hawking modes. A large shift of this kind has a noticeable effect on the outgoing radiation, and indeed in 2-d models [14] the shift is responsible for the Hawking radiation in the usual semi-classical formalism. However, when considering the consequences of small fluctuations in the infalling matter, the shift is unimportant since Hawking radiation is approximately in the Kruskal vacuum close to the horizon. The transformation $x^- \rightarrow a^-(x^- + \Delta^-)$ is thus an approximate symmetry of the Hawking state and the only effect on the radiation as observed at infinity is a small shift in its temperature [8].

More recently, in Refs. [9, 10] it was pointed out that large quantum gravitational effects are observed in the propagation of matter fields falling into the horizon. In Ref. [10] it was shown that an inherent Planck scale uncertainty in the mass of a black hole (due to quantum fluctuations in the infalling matter that forms the black hole) leads to a large uncertainty in the state of matter falling through the horizon on a particular set of spacelike hypersurfaces which we shall refer to as S-surfaces. These are surfaces that capture some of the Hawking radiation at future null infinity, avoid the last stages of evaporation of the hole and then stay outside the horizon in the region above the shock-wave (see Fig. 1). An important element in this calculation is a separation between the matter that forms the hole and that which falls through the horizon after the hole is formed. It is found that a small fluctuation in the black hole mass leads to a shift in both the in- and out-going modes. In a region of the S-surface close to the horizon but not too close to the past horizon, the change in the modes can be expressed in Kruskal coordinates as $x^- \rightarrow a^-(x^- + \Delta^-)$ and $x^+ \rightarrow a^+(x^+ + \Delta^+)$, where Δ^\pm are proportional to the shift in the mass, but are also highly sensitive to the properties of the S-surface. The shift in x^- is small for small fluctuations in mass, so that by the arguments given above there is a negligible effect on the outgoing radiation in this first approximation. However, since the infalling matter is typically found in a state close to the Schwarzschild vacuum, the transformation in x^+ has a non-trivial effect. For a fluctuation as small as $\Delta M \sim e^{-M}$ in Planck units, the shift in x^+ and the size of the region for which it is valid are large enough to change the incoming state appreciably.

In Sec. 2.1 we present a review of the derivation of the shifts $x^- \rightarrow a^-(x^- + \Delta^-)$ and $x^+ \rightarrow a^+(x^+ + \Delta^+)$ whose origin is the pointwise identification of the same spacelike hypersurface in two different classical spacetimes according to its intrinsic geometry. In Sec. 2.1 the shift is derived for CGHS [14] solutions with different masses. This is generalised in Sec. 2.2 to the identification of spherically symmetric spacelike hypersurfaces in Schwarzschild spacetimes with different masses.

In Sec. 3 the effect of the shift on the propagation of matter is discussed. Although

in this paper we will talk mainly about the CGHS model, the results given in Sec. 2.2 show that similar results hold for matter in the S-wave on a Schwarzschild background. The range of quantum fluctuations in the mass of a black hole is determined by the finite lifetime and size of the hole, implying an uncertainty in the mass M of at least $1/M$. These fluctuations are extremely small and should not be observable, so we can consider the effect of tracing over them. The result is a density matrix of the form

$$\rho(\Sigma) = \int dM |\omega(M)|^2 |f_M(\Sigma)\rangle \langle f_M(\Sigma)| \quad (1)$$

where $\omega(M)$ is some weight function peaked around M_0 with a width of order 1, and each $|f_M(\Sigma)\rangle$ is obtained by propagating the matter state from the vacuum state at past null infinity to the S-surface Σ on a black hole background of mass M .

A semiclassical interpretation of this density matrix requires that one define a mean background spacetime (say that with mass M_0)¹, on which ρ evolves, and that one regard each $|f_M(\Sigma)\rangle$ with $M \neq M_0$ as an element of the Hilbert space $\mathcal{H}_{M_0, \Sigma}$ of states on Σ in the spacetime \mathcal{M}_0 with mass M_0 . This can be achieved by using the canonical inner product between states defined on the same spacelike hypersurface embedded in different classical spacetimes [10].

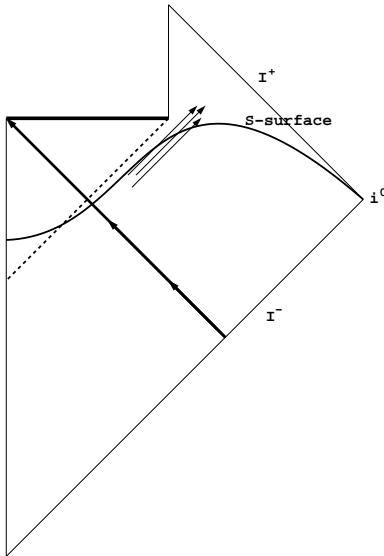


Figure 1: An S-surface shown in an evaporating black hole spacetime. The surface remains outside the event horizon in the region above the shock-wave. The arrows represent the outgoing Hawking radiation.

The semiclassical approximation as it is generally used, assumes that all the $|f_M(\Sigma)\rangle$ are approximately equal. In that case ρ is a pure state density matrix that evolves

¹Note that the choice of mean mass M_0 is somewhat arbitrary since we could choose any mass M within the peak of the weight function $\omega(M)$

according to the functional Schrödinger equation on any background spacetime with a mass close to the mean black hole mass M_0 . This is generally true for generic hypersurfaces Σ . However, on a set of S-surfaces the shift $x^+ \rightarrow a^+(x^+ + \Delta^+)$ causes a change in the $|f_M(\Sigma)\rangle$ that is very sensitive to the value of M . Then ρ is a mixed state reflecting the entanglement of the matter with the quantum gravitational fluctuations. The change in the $|f_M(\Sigma)\rangle$ is concentrated in an extended region of the S-surface adjacent to the horizon, and is large enough that a pair of states $|f_M(\Sigma)\rangle$ and $|f_{\bar{M}}(\Sigma)\rangle$ are approximately orthogonal if $M - \bar{M} \gtrsim e^{-M}$ [10]. This can be taken to indicate a breakdown in the semiclassical approximation.

In Sec. 3 it is shown how each element $|f_M(\Sigma)\rangle$ defines a Fock state on \mathcal{M}_0 :

$$\mathcal{U}(\Sigma, M \rightarrow M_0)|f_{M_0}\rangle \tag{2}$$

which is defined to be equal to $|f_M(\Sigma)\rangle$ when evaluated on Σ in \mathcal{M}_0 . Here $|f_{M_0}\rangle$ is a Fock state that agrees with $|f_M(\mathcal{I}^-)\rangle$ when evaluated at past null infinity. The operator \mathcal{U} implements the shift $x^+ \rightarrow a^+(x^+ + \Delta^+)$ and is constructed explicitly in Sec. 4. It is shown to bear a close resemblance to the operator defined in Ref. [11] to describe the Planckian interactions between in and outgoing matter at the horizon.

The inner product alone does not give any sense of how the shift $x^+ \rightarrow a^+(x^+ + \Delta^+)$ changes the picture of black hole evaporation, since orthogonal states can give rise to very similar physics. In Sec. 5 we present a simple calculation of the energy associated with the orthogonal infalling states, which we shall use to argue both that the breakdown in the semiclassical approximation is large, and that a backreaction on the outgoing Hawking radiation is produced. We shall compute the energy flux of the orthogonal states on an entire S-surface including the region close to the infalling matter that formed the black hole. The energy flux is zero below the shock wave and far above it, but in between it becomes large enough to affect the outgoing Hawking radiation at its early stages.

In Sec. 6 we define a region, determined by the S-surfaces and by the location within the S-surfaces of the large energy fluctuations, in which there is an irrevocable breakdown in the semiclassical approximation. This region is shown to be bounded by a surface approximately one Planck distance from the horizon, resembling a stretched horizon. This stretched horizon represents a boundary beyond which we cannot safely perform calculations armed only with a knowledge of semiclassical physics. An important aspect of our approach, however, is that this breakdown is only manifested when working in a very specific foliation close to the horizon, and a different choice of foliation allows a semiclassical description of horizon physics. We interpret this as a manifestation of black hole complementarity [7], in the sense that different observers have different perceptions of the effects of spacetime fluctuations. For observers that stay outside the horizon and detect some of the Hawking radiation, quantum gravitational effects are large, and there is a strong interaction between infalling matter and the outgoing radiation.

2 A Large Shift

2.1 The CGHS black hole

In this subsection we briefly review the geometrical origin of the large shift in Kruskal coordinates [10] for the CGHS black hole [14]. This shift comes from the identification, through its intrinsic geometry, of a single spacelike hypersurface Σ embedded in different classical spacetimes. The main result we rederive in this subsection is given in Eqs. (8) and (9).

We shall work in the CGHS spacetime [14],

$$\begin{aligned} ds^2 &= -e^{2\rho} dx^+ dx^-, \\ e^{-2\rho} &= e^{-2\phi} = -\frac{M}{\lambda}(\lambda x^+ - 1)\Theta(x^+ - 1/\lambda) - \lambda^2 x^+ x^- \end{aligned} \quad (3)$$

where ϕ is a dilaton field that can be taken to represent the radial component of a four-dimensional metric after dimensional reduction [15]. Eq. (3) represents a black hole formed by the collapse of a shock wave of matter. The x^\pm are referred to as Kruskal coordinates.

The intrinsic geometry of a spacelike hypersurface in the CGHS model can be defined by the function $\phi(s)$ where s is the proper distance to some fixed point on the hypersurface determined by ρ . The fixed point should be defined in a canonical fashion. It is natural to consider the fixed point to be at infinity, since this corresponds to anchoring the surface at infinity, thereby ensuring the validity of the semiclassical approximation there. Two hypersurfaces embedded in different spacetimes have the same intrinsic geometry if they define the same function $\phi(s)$.

Consider two different solutions (3) with masses M and \bar{M} . We can define a spacelike hypersurface Σ in the first spacetime through a relation between x^+ and x^- , say $\lambda x^- = g(\lambda x^+) - M/\lambda$ (where this definition is made for convenience as the horizon is at $\lambda x^- = -M/\lambda$). Given (3), it is easy to deduce the function $\phi(s)$ which defines its intrinsic geometry. Consider also a hypersurface $\bar{\Sigma}$ in the mass \bar{M} spacetime, defined by the relation $\lambda \bar{x}^- = \bar{g}(\lambda \bar{x}^+) - \bar{M}/\lambda$. When Σ and $\bar{\Sigma}$ define the same $\phi(s)$ in two different spacetimes, there is a canonical map between points on the hypersurface in each of the spacetimes which follows from identifying local intrinsic geometry. The condition that $\phi(s)$ is the same for Σ and $\bar{\Sigma}$ is equivalent to the two conditions:

$$\frac{M}{\lambda} - \lambda x^+ g(\lambda x^+) = \frac{\bar{M}}{\lambda} - \lambda \bar{x}^+ \bar{g}(\lambda \bar{x}^+) \quad (4)$$

$$\frac{g(\lambda x^+) + \lambda x^+ g'(\lambda x^+)}{\sqrt{-g'(\lambda x^+)}} = \frac{\bar{g}(\lambda \bar{x}^+) + \lambda \bar{x}^+ \bar{g}'(\lambda \bar{x}^+)}{\sqrt{-\bar{g}'(\lambda \bar{x}^+)}} \quad (5)$$

This set of equations can be solved for $\bar{g}(\lambda \bar{x}^+)$ given $g(\lambda x^+)$. The identification of points with the same local intrinsic geometry is given by functions $\bar{x}^+(x^+)$ and $\bar{x}^-(x^-)$. These functions are essential for comparing states evolved to Σ and $\bar{\Sigma}$ in the two different spacetimes.

The particular class of surfaces that we focus on in this paper, referred to in the introduction as S-surfaces, can be specified by

$$g(\lambda x^+) = -\alpha^2(\lambda x^+ - 1) - \delta \quad (6)$$

The two constants, the slope α , and the distance δ in x^- coordinates between the horizon (at $\lambda x^- = -M/\lambda$) and the point where the surface crosses the shock wave (at $\lambda x^+ = 1$), are chosen so that $\alpha^2 \ll 1$ and $\delta \ll 1$, ensuring that the surface is almost lightlike and runs very close to the horizon.

A particularly simple choice of $g(\lambda x^+)$ is

$$\begin{aligned} \lambda x^- &= -\alpha^2 \lambda x^+ - 2\alpha \sqrt{\frac{M}{\lambda}} - \frac{M}{\lambda} & \lambda x^+ \geq 1 \\ \lambda x^- &= -\left(\alpha + \sqrt{\frac{M}{\lambda}}\right)^2 \lambda x^+ & \lambda x^+ \leq 1 \end{aligned} \quad (7)$$

For this choice of surfaces the scale of α is fixed by the condition that the S-surface capture a proportion r of the Hawking radiation, which requires that $\alpha \sim e^{-rM/\lambda}$.

For (7) we find the following relations $\bar{x}^+(x^+)$. For $\bar{M} < M$:

$$\begin{aligned} \lambda \bar{x}^+ &= \lambda x^+ + \frac{1}{\alpha} \left(\sqrt{\frac{M}{\lambda}} - \sqrt{\frac{\bar{M}}{\lambda}} \right) & \lambda x^+ \geq 1 \\ \lambda \bar{x}^+ &= \left(1 + \frac{1}{\alpha} \sqrt{\frac{M}{\lambda}} \right) \lambda x^+ - \frac{1}{\alpha} \sqrt{\frac{M}{\lambda}} & 1 \geq \lambda x^+ \geq \lambda x^+ < \\ \lambda \bar{x}^+ &= \left(\frac{\alpha + \sqrt{M/\lambda}}{\alpha + \sqrt{\bar{M}/\lambda}} \right) \lambda x^+ & \lambda x^+ < \lambda x^+ \geq 0 \end{aligned} \quad (8)$$

where $\lambda x^+ < (\alpha + \sqrt{\bar{M}/\lambda})/(\alpha + \sqrt{M/\lambda})$. For $\bar{M} > M$:

$$\begin{aligned} \lambda \bar{x}^+ &= \lambda x^+ - \frac{1}{\alpha} \left(\sqrt{\frac{\bar{M}}{\lambda}} - \sqrt{\frac{M}{\lambda}} \right) & \lambda x^+ \geq \lambda x^+ > \\ \lambda \bar{x}^+ &= \left(1 + \frac{1}{\alpha} \sqrt{\frac{\bar{M}}{\lambda}} \right)^{-1} \left(\lambda x^+ + \frac{1}{\alpha} \sqrt{\frac{M}{\lambda}} \right) & \lambda x^+ > \lambda x^+ \geq 1 \\ \lambda \bar{x}^+ &= \left(\frac{\alpha + \sqrt{M/\lambda}}{\alpha + \sqrt{\bar{M}/\lambda}} \right) \lambda x^+ & 1 \geq \lambda x^+ \geq 0 \end{aligned} \quad (9)$$

where $\lambda x^+ > 1 + (\sqrt{\bar{M}/\lambda} - \sqrt{M/\lambda})/\alpha$. The above coordinate relationship translates directly to a relationship between coordinates $\lambda v = \ln(\lambda x^+)$ and $\lambda \bar{v} = \ln(\lambda \bar{x}^+)$ that are asymptotically flat on \mathcal{I}^+ . The relations $\bar{x}^-(x^-)$ can be immediately derived from Eqs. (8) and (9) given expressions for $\lambda x^- = g(\lambda x^+) - M/\lambda$ and $\lambda \bar{x}^- = \bar{g}(\lambda \bar{x}^+) - \bar{M}/\lambda$.

It should be apparent from (8) and (9) that a large effect comes in each case from the first relation, for which $\bar{x}^+ \approx x^+ + \Delta^+$ with Δ^+ of order $1/\alpha$. This relation is valid

for the region of the S-surface above the shock wave, all the way out to i_0 . A large effect also comes from the second relation. Although we might be suspicious of the details of the metric in the neighbourhood of the shock wave of the CGHS solution, and so mistrust this second relation, it is inevitable that the relation involve large parameters to compensate for the shift in the first relation, since the third relation is in each case approximately the identity.

2.2 The Schwarzschild black hole

In this subsection we derive the large shift in Kruskal coordinates for the case of a Schwarzschild black hole following the derivation given in the previous subsection for the CGHS solution. The main result of this subsection is given in Eq. (22).

Taking $G = c = 1$, the Schwarzschild metric can be written as

$$ds^2 = -\frac{2M}{r}e^{-r/2M}dx^+dx^- + r^2d\Omega^2 \quad (10)$$

Here $\kappa x^+ = e^{\kappa v}$ and $\kappa x^- = -e^{-\kappa u}$ where $\kappa = 1/4M$ is the surface gravity, $u = t - r^*$, $v = t + r^*$, $r^* = r + 2M \ln |r/2M - 1|$ is the tortoise coordinate and r is the usual Schwarzschild coordinate. We shall refer to x^+ and x^- as Kruskal coordinates. It follows that

$$-\kappa x^+ \kappa x^- = (r/2M - 1)e^{r/2M}. \quad (11)$$

For the purposes of this paper, we are only interested in the metric near the future horizon, so we expand (10) around $r = 2M$ by writing $r = 2M(1 + R)$, where $0 < R \lesssim M_{pl}/M$, so that $\kappa x^+ \kappa x^- = -eR$.

If we write

$$ds^2 = -e^{2\rho}dx^+dx^- + Ad\Omega^2 \quad (12)$$

then it follows that near the future horizon (*i.e.* when $-\kappa x^+ \kappa x^- \lesssim eM_{pl}/M$ and $\kappa x^+ > 1$),

$$e^{-2\rho} = e - 2\kappa x^+ \kappa x^-, \quad A = \frac{4M^2}{e} (e - 2\kappa x^+ \kappa x^-) \quad (13)$$

This is similar to the CGHS metric in the coordinates of (3) since the dilaton field $e^{-2\phi}$ is the coordinate scalar that represents the area of 2-spheres A after dimensional reduction. It is convenient from now on to set $\hbar = 1$ and work with a dimensionless M defined in Planck units so that $M \gg 1$.

Let us now consider the matching of spacelike hypersurfaces in Schwarzschild solutions with different masses. As a simplification, we shall restrict observations of the matter fields to spherically symmetric spacelike hypersurfaces. Such hypersurfaces are defined by their intrinsic geometry in the t, r plane. The area A of a 2-surface of constant r is a scalar under a coordinate transformation in the t, r plane. The intrinsic geometry of a spherically symmetric hypersurface is thus given by the function $A(s)$ where s is the proper distance from infinity in the t, r plane, determined by ρ . Two hypersurfaces embedded in different spacetimes have the same intrinsic geometry if they define the same function $A(s)$.

As before, we now consider two black holes with masses M and \bar{M} . We can define a spherically symmetric spacelike hypersurface Σ in the first spacetime through a relation between x^+ and x^- , say $x^- = g(\kappa x^+)$ (note that in (10) the horizon is at $x^- = 0$). Given the form of the metric, it is easy to deduce the function $A(s)$ which defines the intrinsic geometry. We then seek a hypersurface $\bar{\Sigma}$ in the mass \bar{M} spacetime, defined by the relation $\bar{x}^- = \bar{g}(\bar{\kappa} \bar{x}^+)$ defining the same function $A(s)$. This is equivalent to the two conditions:

$$A(\kappa x^+, g(\kappa x^+)) = \bar{A}(\bar{\kappa} \bar{x}^+, \bar{g}(\bar{\kappa} \bar{x}^+)) \quad (14)$$

$$\frac{dA}{ds}(\kappa x^+, g(\kappa x^+)) = \frac{d\bar{A}}{d\bar{s}}(\bar{\kappa} \bar{x}^+, \bar{g}(\bar{\kappa} \bar{x}^+)) \quad (15)$$

Plugging (13) into (14) and (15), one gets

$$M^2 (e - 2\kappa x^+ \kappa g) = \bar{M}^2 (e - 2\bar{\kappa} \bar{x}^+ \bar{\kappa} \bar{g}). \quad (16)$$

$$\frac{\kappa g + \kappa x^+ \kappa g'}{\sqrt{-\kappa g'}} = \frac{\bar{\kappa} \bar{g} + \bar{\kappa} \bar{x}^+ \bar{\kappa} \bar{g}'}{\sqrt{-\bar{\kappa} \bar{g}'}}. \quad (17)$$

Here a prime denotes differentiation with respect to κx^+ or $\bar{\kappa} \bar{x}^+$. Differentiating equation (16) with respect to κx^+ and comparing with equation (17) we find

$$\frac{d(\bar{\kappa} \bar{x}^+)}{d(\kappa x^+)} = \frac{M^2 \sqrt{-\kappa g'}}{M^2 \sqrt{-\bar{\kappa} \bar{g}'}}. \quad (18)$$

Equation (18) can be integrated using equations (16) and (17) to give :

$$\ln(\bar{\kappa} \bar{x}^+) = \frac{2M^2}{\bar{M}^2} \int \frac{d(\kappa x^+) \kappa g'}{(\kappa g + \kappa x^+ \kappa g') \mp \sqrt{(\kappa g + \kappa x^+ \kappa g')^2 - 4M^2 \kappa g' (\epsilon + \kappa x^+ \kappa g) / \bar{M}^2}}, \quad (19)$$

where $\epsilon \approx e\Delta M/M$, and $\Delta M = \bar{M} - M$.

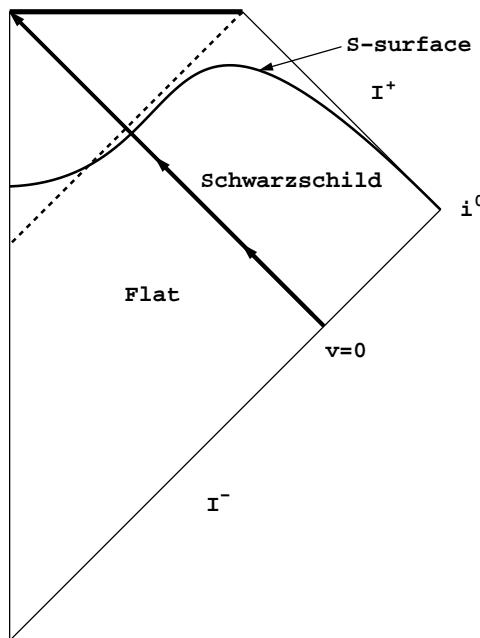


Figure 2: An S-surface in a Schwarzschild spacetime formed by a collapsing shell of matter at $v = 0$. The horizon is at $x^- = 0$.

As in the previous subsection, an S-surface can be specified by

$$\kappa g(\kappa x^+) = -\alpha^2 \kappa x^+ - \delta \quad (20)$$

where the two constants α and δ are chosen so that $\alpha \ll \delta \ll 1$, ensuring that the surface is almost lightlike, and runs very close to the horizon. It follows that in the region of validity of our approximation, $\kappa x^+ \kappa x^- \lesssim e/M$ and $\kappa x^+ > 1$ in which (19) is valid, κx^- effectively takes the constant value $-\delta$. Under the same condition, we can easily evaluate (19) for $e/\Delta M \gtrsim \kappa x^+ > 1$ so that

$$\ln(\bar{\kappa} \bar{x}^+) = \ln\left(\kappa x^+ - \frac{\epsilon}{\delta}\right) + C \quad (21)$$

where C is a constant of integration. Using $\kappa x^+ = e^{\kappa v}$, the above coordinate relationship translates to a coordinate relationship between the null tortoise coordinates (that are asymptotically flat on \mathcal{I}^+) of the form

$$\bar{\kappa} \bar{v} = \ln\left(e^{\kappa v} - \frac{e\Delta M}{M\delta}\right) + C. \quad (22)$$

This is the main result of this subsection, which we emphasize is valid on the region of the S-surface above the collapsing shell and close to the future horizon (see Fig. 2). A discussion of the shift for general black hole spacetimes will be given in Ref. [16].

3 An Effective semiclassical Theory

In this section we turn to the propagation of a matter state on a quantised gravitational background. A state representing a black hole spacetime of mass M must, as discussed in the introduction, have an uncertainty in its mass which is at the very least of the order of $1/M$ in Planck units (we shall, for convenience, take it to be 1). As discussed in [17], a state for the gravitational field that is localised around a particular classical black hole solution with mass M_0 , can be regarded as a superposition of plane wave WKB states, each of which is an approximate mass eigenstate:

$$|\Psi_{\text{gravity}}\rangle = \int dM \omega(M) |M\rangle \quad \text{or} \quad \Psi[h_{ij}] = \int dM \omega(M) \Psi_M[h_{ij}] \quad (23)$$

Here $\omega(M)$ is a weight function that is centered around a mean value M_0 with a spread of order 1. We can approximate the propagation of matter on a *quantum* black hole background by looking at propagation on classical backgrounds with different masses in the neighbourhood of M_0 , so that the combined matter–gravity state is of the form

$$|\Psi(\Sigma)\rangle = \int_{M_0-1}^{M_0+1} dM \omega(M) |M\rangle |f_M(\Sigma)\rangle \quad (24)$$

where $|f_M(\Sigma)\rangle$ is a Schrödinger picture state propagating on a background spacetime \mathcal{M} with mass M [18].

It is natural to trace over the gravitational degrees of freedom in (24) since fluctuations on the Planck scale are unobservable. However, in normal semiclassical evolution, the propagation of matter fields should be insensitive to such small fluctuations in the mass of the background, so that all the $|f_M(\Sigma)\rangle$ for different M are approximately equal. In that case there is no entanglement between the matter and gravity degrees of freedom and tracing leads to a pure state density matrix, which is what we observe in everyday experiments. However, we shall show that this is not always the case.

The simply seemingly simple statement of the previous paragraph that the $|f_M(\Sigma)\rangle$ for different M are approximately equal raises the question of how we should compare matter states $|f_M(\Sigma)\rangle$ propagating on different spacetimes with different values of M . In canonical quantum gravity, a state

$$\Psi[h_{ij}, f] \tag{25}$$

takes the form of a correlation between a matter configuration and a spacelike hypersurface (h_{ij}) without making reference to a *spacetime*. Correlations between matter configurations and spacelike hypersurfaces are also contained in the Schrödinger evolution of a quantum matter field on any classical spacetime. These correlations should be regarded as fundamental from the point of view of quantum gravity. Through them a comparison of states defined on different spacetimes, but on the same hypersurface (same h_{ij}) can be made. It is for this reason that the embedding of a hypersurface in different spacetimes was studied in the previous section.

A detailed discussion of the comparison of matter states defined on neighbouring background spacetimes can be found in [10] and [17]. The essential idea is the existence of a canonical Hilbert space associated with a spacelike hypersurface, independent of any spacetime. The value of a good old-fashioned Schrödinger picture state on a particular hypersurface can be expressed as a state in this canonical Hilbert space. Comparisons between Schrödinger picture states defined on different spacetimes then also take place through this Hilbert space.

The Hilbert space associated with a surface Σ is defined in terms of a set of geometric modes on Σ (for example of the form $e^{i\omega s}$ where s is the proper distance on Σ_0). At least in 2D, there exists a canonical isomorphism between the Fock space defined with these modes and the Hilbert space of all Schrödinger picture states, which comes from evaluating the Schrödinger picture state on Σ in an appropriate foliation [10].

At past null infinity, we can consider a null hypersurface of constant u or x^- , which we refer to as \mathcal{I}^- . Since everything is flat, the Hilbert space intrinsic to \mathcal{I}^- is isomorphic to the Hilbert space of Schrödinger picture states evaluated at \mathcal{I}^- in an obvious way. Through the Hilbert space on \mathcal{I}^- it is therefore straightforward to say that states defined on different spacetimes are the same at \mathcal{I}^- – they have the same Fock space decomposition with respect to plane wave modes at past null infinity. When comparing states on later-time hypersurfaces, we take the boundary condition that all states are equal in this (intuitive and formal) sense at \mathcal{I}^- .

On later-time hypersurfaces, the comparison through the intrinsic Hilbert space shows that states propagated on different spacetimes are no longer identical. However, if the difference in mass between the spacetimes is small, then for all practical purposes, the states are equal for almost any choice Σ . Under such circumstances, (24) takes the form

$$|\Psi(\Sigma)\rangle \approx \left(\int dM \omega(M) |M\rangle \right) |f_{M_0}(\Sigma)\rangle \quad (26)$$

(where M_0 is a chosen mean value for M) and the semiclassical approximation [18] of treating matter as propagating on a single background spacetime is extremely good.

We can use (24) to make sense of a trace over the unobserved fluctuations in the gravitational background if we work in the Schrödinger picture. For a state of the form (26), tracing over the $|M\rangle$ leaves a pure state density matrix

$$\rho_\Psi(\Sigma) = |f_{M_0}(\Sigma)\rangle \langle f_{M_0}(\Sigma)| \quad (27)$$

In general, however, tracing (24) gives

$$\rho_\Psi(\Sigma) = \int dM |\omega(M)|^2 |f_M(\Sigma)\rangle \langle f_M(\Sigma)| \quad (28)$$

where one should think of all states $|f_M(\Sigma)\rangle$ as being in the Hilbert space intrinsic to Σ . The density matrix ρ can be considered the effective state of matter with the gravitational fluctuations integrated out, at least when the departure from semiclassical behaviour is not too large. Working with ρ we can evaluate the expectation value of operators in the usual way:

$$\langle \mathcal{O} \rangle = \text{Tr}(\rho \mathcal{O}) \quad (29)$$

For simplicity, let us assume that $\omega(M)$ is roughly constant over some range $M_0 - \Delta M_{max} \leq M \leq M_0 + \Delta M_{max}$ centered around a mass M_0 , with ΔM_{max} of order 1, and is zero outside this range. The semiclassical approximation requires that states $|f_M(\Sigma)\rangle$ of (24) be approximately equal. We can test this by evaluating the inner product between the elements $|f_M(\Sigma)\rangle$ for a pair of values of M within the specified range. It was found in [10] that this inner product is approximately zero unless the difference between the values of M differ at most on a scale of the order of e^{-M} .

The breakdown in the semiclassical approximation implied by this result can be expressed in terms of the entropy of the density matrix (28). The entropy is zero for a pure state (27), but becomes non-zero as the overlap

$$\langle f_M(\Sigma) | f_{\bar{M}}(\Sigma) \rangle \quad (30)$$

between states in the density matrix decreases. An estimate of the entropy of this density matrix, assuming $\omega(M)$ to be of the form outlined above, comes from looking at the number of approximately orthogonal states that are contained in ρ . If for a given surface

$$|\langle f_M(\Sigma) | f_{\bar{M}}(\Sigma) \rangle|^2 \ll 1 \quad (31)$$

for a ΔM much smaller than 1 (the characteristic size of the fluctuations in M), then we can approximate ρ by changing the integral in (28) to a sum over states that are approximately orthogonal:

$$\rho_\Psi(\Sigma) = \frac{1}{N} \sum_{l=-N/2}^{N/2} |f_{M_0+l\Delta M_0}(\Sigma)\rangle\langle f_{M_0+l\Delta M_0}(\Sigma)| \quad (32)$$

where ΔM_0 is the value of $M - M_0$ required for the overlap (30) to be effectively zero, and $N = \Delta M_{max}/\Delta M_0$, where ΔM_{max} defines the size of the fluctuations in M and is of order 1. This gives an estimate of the entropy $\text{Tr}(\rho \ln \rho) \sim \ln N$ of ρ . This entropy [9, 10] can be interpreted as the entanglement entropy between the gravitational and matter degrees of freedom, and we refer to it as the gravitational entropy associated with the surface Σ .

It is useful to define an operator that contains information about the difference between states $|f_M(\Sigma)\rangle$ and $|f_{\bar{M}}(\Sigma)\rangle$ on a hypersurface Σ . The simplest way to define such an operator is to work in a Fock space picture. Any Schrödinger picture state $|f_M(\Sigma)\rangle$ is associated with a unique Fock state $|f_M\rangle$ on the Hilbert space of the spacetime with mass M . Any state $|f_{\bar{M}}(\Sigma)\rangle$ can be associated with a Fock state in the same Hilbert space in a hypersurface dependent way: For a given Σ define a Schrödinger picture state $|g_M(\Sigma)\rangle$ such that it is equal to $|f_{\bar{M}}(\Sigma)\rangle$ in the Hilbert space intrinsic to Σ . Then associate $|f_{\bar{M}}(\Sigma)\rangle$ with the Fock state $|g_M\rangle$ on the Hilbert space of the spacetime with mass M .

We can describe the difference between states $|f_M(\Sigma)\rangle$ and $|f_{\bar{M}}(\Sigma)\rangle$ on a hypersurface Σ by the operator $\mathcal{U}(\Sigma, \bar{M} \rightarrow M)$ so that

$$|f_{\bar{M}}\rangle \rightarrow |g_M\rangle \equiv \mathcal{U}(\Sigma, \bar{M} \rightarrow M)|f_M\rangle \quad (33)$$

Note that the operator is independent of the choice of state $|f_M\rangle$. When $\bar{M} = M$, \mathcal{U} is of course the identity operator for all Σ . For $\bar{M} \neq M$, it implements the coordinate transformations $\bar{x}^+(x^+)$ and $\bar{x}^-(x^-)$ [10] and is constructed explicitly in the next section. \mathcal{U} defined in this way can be thought of as an interaction picture time evolution operator (because of its dependence on Σ) incorporating the quantum gravitational effects on the matter states.

In [10] the overlap (30) was computed for a family of S-surfaces. For these surfaces, \mathcal{U} implements the shifts $x^+ \rightarrow a^+(x^+ + \Delta^+)$ and $x^- \rightarrow a^-(x^- + \Delta^-)$. The right-moving sector of outgoing matter is in the Kruskal vacuum close to the horizon, and since Δ^- is small for a small fluctuation, the shift $x^- \rightarrow a^-(x^- + \Delta^-)$ is an approximate symmetry of the outgoing state. The infalling matter, on the other hand, is in the Schwarzschild vacuum, and so the transformation $x^+ \rightarrow a^+(x^+ + \Delta^+)$ has a non-trivial effect that becomes large as Δ^+ grows. Thus for S-surfaces, for which Δ^+ is large, ρ should represent a mixed state. For this reason we shall focus exclusively on the infalling (or left-moving) matter, and we shall take \mathcal{U} to be the identity operator on the right-moving sector. Eq. (31) is satisfied for very small $\Delta M \ll 1$ for S-surfaces [10], and this can be taken to indicate a breakdown in the semiclassical approximation.

Note that since the large shift is confined to S-surfaces, the entanglement between matter and gravity represented by \mathcal{U} is a phenomenon associated with specific foliations of spacetime [10, 17]. For example, a foliation of the region close to the horizon that is associated with a one-parameter family of α 's and δ 's of Eq. (6) that are not both small, does not give rise to a large shift in the Kruskal coordinates, and \mathcal{U} for such hypersurfaces is trivial. The consequences of this are discussed later.

4 The \mathcal{U} operator

The effect of the operator $\mathcal{U}(\Sigma, \bar{M} \rightarrow M)$ is to transform the vacuum state with respect to modes $\phi_i(x^+)$ to the vacuum with respect to modes $\phi_i(\bar{x}^+(x^+))$. Equivalently, the vacuum with respect to modes $\phi_i(v)$ becomes that with respect to $\phi_i(\bar{v}(v))$ where as before $\lambda v = \ln(\lambda x^+)$ and $\lambda \bar{v} = \ln(\lambda \bar{x}^+)$.

Consider expanding the matter field operator as either

$$\hat{\phi}(v) = \int_0^\infty \frac{d\omega}{\sqrt{2\omega}} (a_\omega e^{-i\omega v} + a_\omega^\dagger e^{i\omega v}) \quad (34)$$

or

$$\hat{\phi}(v) = \int_0^\infty \frac{d\omega}{\sqrt{2\omega}} (b_\omega e^{-i\omega \bar{v}(v)} + b_\omega^\dagger e^{i\omega \bar{v}(v)}) \quad (35)$$

It follows that

$$b_\omega = \int_0^\infty d\omega' (a_{\omega'} \alpha_{\omega'\omega}^* + a_{\omega'}^\dagger \beta_{\omega'\omega}^*) \quad (36)$$

where $\alpha_{\omega\omega'}$ and $\beta_{\omega\omega'}$ are the Bogoliubov coefficients:

$$\alpha_{\omega\omega'} = \frac{1}{2\pi} \sqrt{\frac{\omega'}{\omega}} I_{\omega\omega'}^+ \quad \text{and} \quad \beta_{\omega\omega'} = \frac{1}{2\pi} \sqrt{\frac{\omega'}{\omega}} I_{\omega\omega'}^- \quad (37)$$

and $I_{\omega\omega'}^\pm$ are the integrals [10]

$$I_{\omega\omega'}^\pm = \int_{-\infty}^\infty d\bar{v} e^{-i\omega v(\bar{v}) \pm i\omega' \bar{v}}. \quad (38)$$

The vacuum defined by $a_\omega |0\rangle = 0$ is then related to the state $\mathcal{U}(\Sigma, \bar{M} \rightarrow M) |0\rangle$ defined by $b_\omega \mathcal{U}(\Sigma, \bar{M} \rightarrow M) |0\rangle = 0$ by the operator

$$\mathcal{U}(\Sigma, \bar{M} \rightarrow M) = B * \exp \left[-\frac{1}{2} a_{\omega'}^\dagger D_{\omega'\omega} a_\omega^\dagger \right] \quad (39)$$

where D is the symmetric matrix

$$D_{\omega'\omega} = \int d\omega'' \beta_{\omega''\omega'}^* \alpha_{\omega\omega''}^{-1} \quad (40)$$

and B is a normalization constant which is equal to $\langle 0 | \mathcal{U}(\Sigma, \bar{M} \rightarrow M) | 0 \rangle^{-1}$.

In order to compute the Bogoliubov coefficients $\alpha_{\omega\omega'}$ and $\beta_{\omega\omega'}$, the coordinate relation $\bar{v}(v)$ is needed throughout Σ . However, working in a wave-packet basis the relation $\bar{v}(v)$ in a restricted region is sufficient to approximate the relevant components of the Bogoliubov coefficients. This allows one to isolate local features of the operator \mathcal{U} (in a wave packet basis the index ω will be replaced by the indices (j, n) [19]). For $\lambda v > \Delta M/(\lambda\alpha\sqrt{M/\lambda})$ and $\lambda v < (\alpha + \sqrt{M/\lambda})/(\alpha + \sqrt{M/\lambda})$, the relevant β coefficients are essentially zero. It follows that the state $|0\rangle$ is only affected in a finite region, and that outside of this region \mathcal{U} is effectively the identity operator. For more complicated spacetimes, such as the Schwarzschild solution, we cannot obtain the exact relation across classical infalling matter, and even in the CGHS case, it is not clear that we should trust our results entirely in that region, but a local analysis still enables one to say a considerable amount about \mathcal{U} .

One can see from (39) that the operator \mathcal{U} in a wave packet basis creates “localized particles” in a region where the β Bogoliubov coefficients are non zero. If we restrict our observation to a certain region (say above the shock wave) then we lose the correlations between the regions above and below the shock wave. For example, for $\Delta M < 0$, restricting attention to the region above the shock wave, but below $\lambda v = \Delta M/(\lambda\alpha\sqrt{M/\lambda})$, the particle creation appears thermal. That is

$$\beta_{jn\omega'}^* \approx -e^{-\pi\omega_j/\kappa} \alpha_{jn\omega'}. \quad (41)$$

(here n labels a wave-packet in this region). Note that the temperature defined by this relation is the same as the temperature of the black hole. The thermal form of \mathcal{U} in this region comes about because of a dramatic expansion in the map from the \bar{v} to the v coordinates. This expansion is compensated by an equally dramatic squashing in the map in the region $0 \geq \lambda v \geq \ln(\lambda x_<^+)$ since for $v < \ln(\lambda x_<^+)$ the map is the identity. For $\Delta M > 0$, the reverse occurs, so that the operator \mathcal{U} appears thermal for modes in the interval $\ln(\lambda x_>^+) > v > 0$ where there is an expansion in the coordinate relation, but this is compensated by a region of squashing for $v > \ln(\lambda x^+)$. Similar results apply for the Schwarzschild spacetime (with λ replaced by κ) from the results of Sec. 2.2. The use of the word thermal should not of course be taken to mean that the operator \mathcal{U} is not unitary.

There is another formal expression for the operator \mathcal{U} which is useful. For the region above the shock wave it takes the form

$$\mathcal{U}(\Sigma, \bar{M} \rightarrow M) = \exp \left\{ i \int dx^+ \Delta^+ T_{++}(x^+) \right\} \quad (42)$$

where Δ^+ is the shift defined by (8) or (9). From (42) one can easily see that \mathcal{U} is unitary. It follows from Eqs. (34) and (35) and from the commutation relation

$$[T_{++}(x_1^+), \phi(x_2^+)] = i\delta(x_1^+ - x_2^+) \partial_+ \phi(x_2^+) \quad (43)$$

that

$$b_i = \mathcal{U} a_i \mathcal{U}^{-1}, \quad b_i^\dagger = \mathcal{U} a_i^\dagger \mathcal{U}^{-1}. \quad (44)$$

and thus that $|0_b\rangle = \mathcal{U}|0_a\rangle$ and that (39) and (42) implement the same transformation.

We can now make contact with Ref. [11]. There an analogous operator \mathcal{U} is defined, which takes the form

$$\mathcal{U}_{KVV} = \exp \left\{ i \int \int dx^+ dx^- T_{++}(x^+) T_{--}(x^-) \right\} \quad (45)$$

The effect of this operator on the infalling matter is given by treating T_{--} as a classical perturbation on the background. With this interpretation, (42) and (45) are identical in the region $x^+ > 1$ for $\Delta M < 0$ (or $x^+ > x_>^+$ for $\Delta M > 0$). This can be seen by noticing that $\Delta^+ \approx e^{\kappa u_0}$ where u_0 is the value of the retarded time characterising an S-surface. If we take the energy-momentum tensor of the outgoing radiation in [11] to be $T_{uu}(u) = \Delta M \delta(u - u_0)$, the correspondence becomes clear.

This correspondence arises despite the fact that the fluctuations in the background are in our case due to fluctuations in the earlier infalling matter and in Ref. [11] due to fluctuations due to outgoing Hawking radiation. The essential element of either calculation is that these fluctuations are hugely amplified in their effect on the infalling state close to the horizon².

5 Induced Energy-Momentum Tensor

In this section we compute the expectation value of the energy-momentum tensor of a state $\mathcal{U}(\Sigma, \bar{M} \rightarrow M)|0\rangle$ when Σ is an S-surface. We have seen that on an S-surface the coordinate relationship between the identified points on space times with mass M and mass \bar{M} takes the form in the Kruskal coordinates

$$\bar{x}^+ = a^+(x^+ + \Delta^+) \quad (46)$$

and that the operator \mathcal{U} implements a map between modes $\phi(v)$ and modes $\phi(\bar{v}(v))$. It follows from a standard calculation that [4]:

$$\langle T_{vv}^{\bar{M}}(\Sigma) \rangle \equiv \langle 0 | \mathcal{U}^{-1}(\Sigma, \bar{M} \rightarrow M) T_{vv} \mathcal{U}(\Sigma, \bar{M} \rightarrow M) | 0 \rangle \quad (47)$$

$$= \frac{1}{12\pi} (\partial_v \bar{v})^{1/2} \partial_v^2 (\partial_v \bar{v})^{-1/2} \quad (48)$$

and all other components are zero. Since T_{vv} is conserved, this quantity can be interpreted as the energy density of the state $\mathcal{U}(\Sigma, \bar{M} \rightarrow M)|0\rangle$ at \mathcal{I}^- if it were propagated back freely on M , where it would correspond to a distribution of infalling matter.

The simple expression (48) allows us to compute the energy-momentum tensor associated with the state $\mathcal{U}(\Sigma, \bar{M} \rightarrow M)|0\rangle$ in any local region with the knowledge of

²Note, however, that in our treatment we regard \mathcal{U} as having changed the *in* vacuum (the Schrödinger picture), whereas in [11] \mathcal{U} changes the operators rather than the states (the Heisenberg picture).

the coordinate transformation $\bar{v}(v)$. Assuming that $\bar{v}(v)$ has at least a continuous third derivative one can integrate by parts so that

$$\int_{x_1}^{x_2} dv \langle T_{vv}^{\bar{M}}(\Sigma) \rangle = \frac{1}{12} \left[(\partial_v \bar{v})^{1/2} \partial_v (\partial_v \bar{v})^{-1/2} \right] \Big|_{x_1}^{x_2} + \frac{1}{48} \int_{x_1}^{x_2} dv \left(\frac{\partial_v^2 \bar{v}}{\partial_v \bar{v}} \right)^2 \quad (49)$$

Since the coordinate relationship is such that $\bar{v}(v) \rightarrow v$ when $v \rightarrow \pm\infty$ (below the shock wave or at \mathcal{I}^+), the total energy flux

$$E_{tot}^{\bar{M}}(\Sigma) \equiv \int_{-\infty}^{\infty} dv \langle T_{vv}^{\bar{M}} \rangle = \frac{1}{48\pi} \int_{-\infty}^{\infty} \left(\frac{\partial_v^2 \bar{v}}{\partial_v \bar{v}} \right)^2 > 0. \quad (50)$$

of the state must be positive.

Eqs. (8) and (9) give explicit expressions for $\bar{v}(v)$ for the CGHS model when Σ is an S-surface. In the three regions of (8) and (9), the function $\bar{v}(v)$ has the form

$$\bar{v}(v) = \frac{1}{\lambda} \ln \left[e^{\lambda v} + \Delta^+ \right] + \frac{1}{\lambda} \ln a. \quad (51)$$

However, because of the shock wave in the CGHS solution, the relation $\bar{v}(v)$ is continuous but not differentiable. With a smooth classical infalling matter distribution, the function $\bar{v}(v)$ would be smooth. Unfortunately it is much more difficult to obtain exact expressions with a smoothed metric. Exact expressions can still be obtained when the infalling matter arrives in a shock wave of some finite width instead of a delta function, but this still leads to discontinuities in the energy-momentum tensor of the same order of magnitude. We shall therefore restrict our attention to the results obtained with the coordinate relations (8) and (9). In this case we cannot obtain explicit expressions for the total energy flux or for the energy-momentum tensor in the neighbourhood of the singularities caused by the shock wave.

Proceeding with (51) for $\bar{v}(v)$, one gets that

$$\langle T_{vv}^{\bar{M}}(\Sigma) \rangle = \frac{\lambda^2}{48\pi} \left[1 - \frac{1}{(1 + \Delta^+ e^{-\lambda v})^2} \right]. \quad (52)$$

in each of the three regions. From (8) and (9), for $\Delta M < 0$

$$\begin{aligned} \Delta^+ &= \frac{1}{\alpha} \left(\sqrt{\frac{M}{\lambda}} - \sqrt{\frac{\bar{M}}{\lambda}} \right) & \lambda v \geq 0 \\ \Delta^+ &= -\sqrt{\frac{\bar{M}}{\lambda}} \left(\alpha + \sqrt{M/\lambda} \right)^{-1} & 0 \geq \lambda v \geq \ln(\lambda x_<^+) \\ \Delta^+ &= 0 & \ln(\lambda x_<^+) \geq \lambda v \end{aligned} \quad (53)$$

where recall that $\lambda x_<^+ = (\alpha + \sqrt{M/\lambda})/(\alpha + \sqrt{M/\lambda})$. For $\Delta M > 0$,

$$\begin{aligned} \Delta^+ &= -\frac{1}{\alpha} \left(\sqrt{\frac{\bar{M}}{\lambda}} - \sqrt{\frac{M}{\lambda}} \right) & \lambda v \geq \ln(\lambda x_>^+) \\ \Delta^+ &= \frac{1}{\alpha} \sqrt{\frac{M}{\lambda}} & \ln(\lambda x_>^+) \geq \lambda v \geq 0 \\ \Delta^+ &= 0 & 0 \geq \lambda v \end{aligned} \quad (54)$$

where $\lambda x_>^+ = 1 + (\sqrt{\bar{M}/\lambda} - \sqrt{M/\lambda})/\alpha$. These expressions can be used to compute the value of $\langle T_{vv}^{\bar{M}}(\Sigma) \rangle$ in each region. Recall that α is a very small number of the order of e^{-M} in Planck units.

Using (53) and (54), we see that for $\Delta M < 0$, $\langle T_{vv}^{\bar{M}}(\Sigma) \rangle$ is small at large λv , but grows as $\lambda v \rightarrow 0$. The region where $\langle T_{vv}^{\bar{M}}(\Sigma) \rangle$ becomes non-negligible corresponds to the region $\lambda v \lesssim \Delta M/(\lambda\alpha\sqrt{M/\lambda})$ where the β coefficients become non-zero and \mathcal{U} starts to appear thermal in a wave-packet basis. In this region we can derive a similar result for S-wave propagation on a Schwarzschild spacetime using (22). Throughout this region, down to $\lambda v = 0$, $\langle T_{vv}^{\bar{M}}(\Sigma) \rangle \sim \lambda^2/48\pi$. In the small region $0 > \lambda v > \ln(\lambda x_<^+)$, $\langle T_{vv}^{\bar{M}}(\Sigma) \rangle$ is negative and very large, taking the value $-\lambda M/48\pi\alpha^2$ at $\lambda v = \ln(\lambda x_<^+)$. Summing the contributions from these two regions, the total energy flux on the surface appears to be negative, but (50) implies that the singular points must contribute a large energy flux to compensate for the total negative energy flux of (52).

For the case $\Delta M > 0$, $\langle T_{vv}^{\bar{M}}(\Sigma) \rangle$ is again zero for large λv , but becomes large and negative as λv enters the region where the coordinate transformation $\bar{v}(v)$ is non-trivial. The fact that in this case the energy flux is negative can be understood because the transformation $\bar{v}(v)$ squashes a large interval in \bar{v} to a small interval in v . For $\lambda v = \ln(\lambda x_>^+)$, $\langle T_{vv}^{\bar{M}}(\Sigma) \rangle \sim -\lambda^2\Delta M^2/192\pi\alpha^2M$. In the region $\ln(\lambda x_>^+) > \lambda v > 0$ there is a positive energy flux of order $\lambda^2/48\pi$. The total energy flux is negative, so that again the singular points must contribute a large energy flux. Of course as $\Delta M \rightarrow 0$ from above or below the energy-momentum tensor in each region approaches zero.

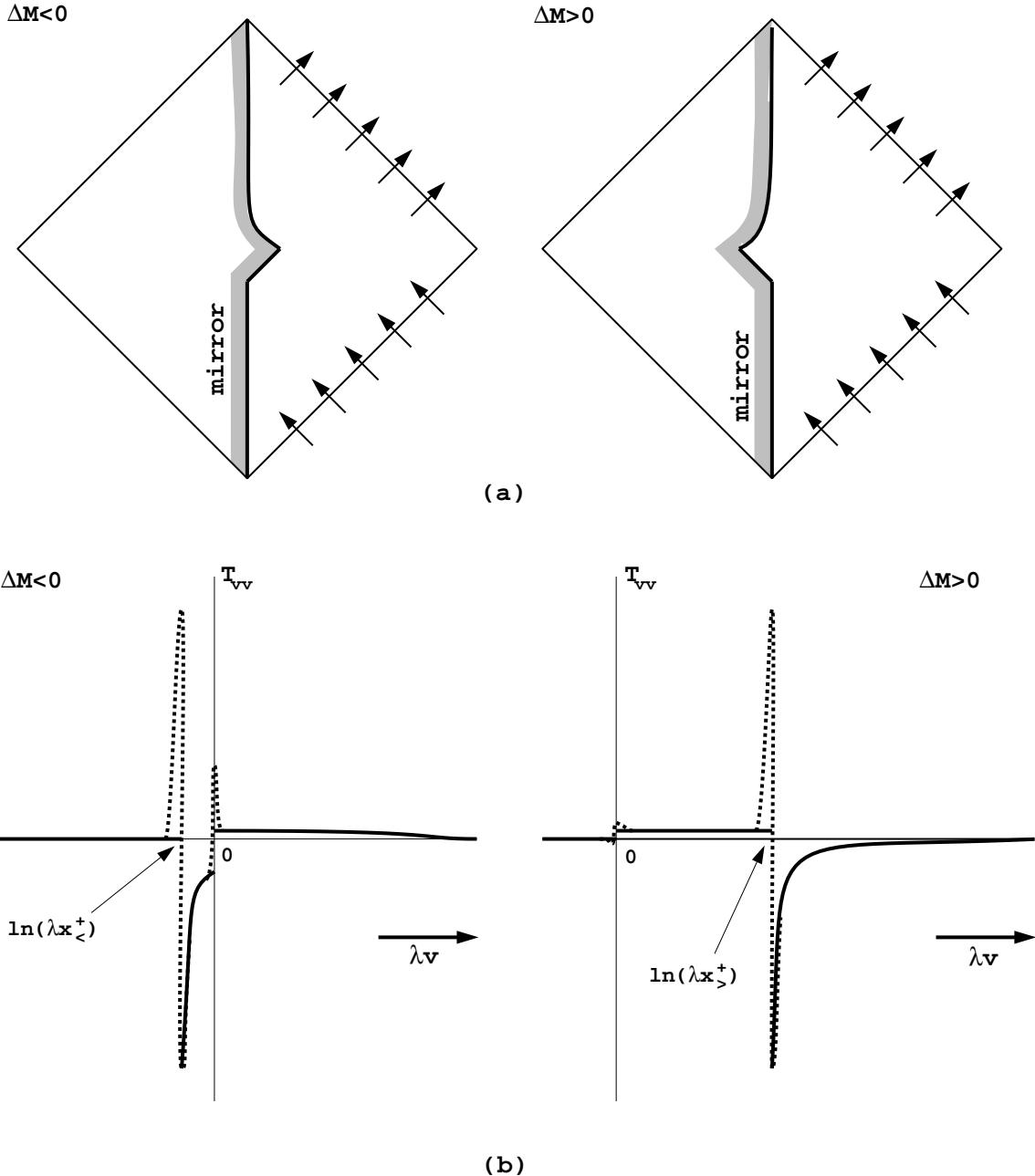


Figure 3: (a) The mirror trajectories. (b) The energy flux distribution for the cases $\Delta M < 0$ and $\Delta M > 0$. The undotted lines represent (52). The dotted lines are an extrapolation in the vicinity of the jumps in the coordinate relations (8) and (9) where some energy flux must be present according to (58).

Eq. (48) for $\langle T_{vv}^{\bar{M}}(\Sigma) \rangle$ is most familiar from the problem of quantum field theory in the presence of a moving mirror. It is useful to translate the coordinate transformations

$\bar{v}(v)$ given by (8) and (9) into the moving mirror language to get an intuitive picture of the generation of energy by the operator \mathcal{U} .

Given a continuous function $\bar{v}(v)$ we can find a mirror trajectory that converts the initial vacuum (at \mathcal{I}^-) to the state $\mathcal{U}(\Sigma, \bar{M} \rightarrow M)|0\rangle$ (at \mathcal{I}^+). The trajectory $x = z(t)$ satisfies

$$\bar{v}(v) - v = 2z \left(\frac{v + \bar{v}(v)}{2} \right) \quad (55)$$

One can solve for the trajectory of the mirror. For the general $\bar{v}(v)$ given by (51) one finds that

$$2\lambda z(t) = \ln \left\{ \frac{\Delta^+ + \sqrt{(\Delta^+)^2 + 4e^{2\lambda t}/a}}{-\Delta^+ + \sqrt{(\Delta^+)^2 + 4e^{2\lambda t}/a}} \right\} + \ln a \quad (56)$$

In Figure 3 we give a plot of the mirror trajectory and of $\langle T_{vv} \rangle$ associated with each of the trajectories. It is clear from the figure that the mirror trajectory is not smooth. The kinks in the trajectory must contribute enough energy flux to surpass the lower bound given by (58).

We can use (50) itself to estimate the positive energy flux that comes from the singular regions. For the coordinate relationship (51),

$$\partial_v^2 \bar{v} = \lambda \partial_v \bar{v} (1 - \partial_v \bar{v}) \quad (57)$$

Thus a lower bound on the total energy flux is given by

$$E_{tot}^{\bar{M}}(\Sigma) = \frac{\lambda^2}{48\pi} \left[\int_{-\infty}^{v_1} (1 - \partial_v \bar{v})^2 + \int_{v_1}^{v_2} (1 - \partial_v \bar{v})^2 + \int_{v_2}^{\infty} (1 - \partial_v \bar{v})^2 \right] \quad (58)$$

where v_1 and v_2 are the boundaries of the regions for which $\bar{v}(v)$ is smooth. This gives a lower bound of

$$E_{tot}^{\bar{M}}(\Sigma) \gtrsim \frac{\lambda^2}{96\pi} \frac{2|\Delta M/\lambda| \sqrt{M/\lambda}}{\alpha(2\alpha\sqrt{M/\lambda} + |\Delta M/\lambda|)} \quad (59)$$

for $\bar{M} < M$, and of

$$E_{tot}^{\bar{M}}(\Sigma) \gtrsim \frac{\lambda^2}{48\pi} \frac{\Delta M/\lambda}{2\alpha\sqrt{M/\lambda}} \quad (60)$$

for $\bar{M} > M$.

Evaluating the ensemble average of the total energy flux (in the sense of equation (29)) we find that

$$\langle E_{tot}(\Sigma) \rangle \sim \frac{1}{\alpha\sqrt{M/\lambda}} = e^{\lambda u_0} \quad (61)$$

where λu is the retarded time at \mathcal{I}^+ and λu_0 is a particular value depending on the S-surface. It is the value of λu where the S-surface intersects the shock wave and, because of the nature of an S-surface, throughout the region where there is a large induced energy-momentum tensor. Now if the total induced energy flux becomes of the order

of the mass of the black hole we would expect that the semiclassical approximation ceases to be valid. In particular we expect the Hawking radiation to be affected. This gives a characteristic time for the departure of the Hawking radiation from its expected form of

$$u_0 \sim \frac{1}{\lambda} \ln(M/\lambda) \quad (62)$$

for the CGHS black hole. For the Schwarzschild black hole, using the results of Sec. 2.2, we can estimate this time to be

$$u_0^{Sch} \sim M \ln M \quad (63)$$

(remember that M is the black hole mass in Planck units).

The fact that the expectation value of T_{vv} becomes large suggests that outside observers should see very different spacetime dynamics to what one would expect. This might be connected to the ‘bounce’ model discussed in [12].

It is unclear how seriously we should take the precise form of $\langle T_{vv}^{\bar{M}}(\Sigma) \rangle$ that we have computed here. The fact that the absolute value of $\langle T_{vv}^{\bar{M}}(\Sigma) \rangle$ is extremely large for different M and sensitive to ΔM indicates that the semiclassical approximation is breaking down in a dramatic way, since the different states $\mathcal{U}(\Sigma, M \rightarrow M_0)|0\rangle$ in (28) contribute to an expectation value for the operator T_{vv} (in the sense of (29)) that has huge quantum fluctuations. The expectation value for the total energy flux on an S-surface (59) and (60) are clearly so large that the methods used to compute them are no longer consistent. Nevertheless, for both $\Delta M < 0$ and $\Delta M > 0$ we can have confidence in our results in the region of large λv far from the shock wave where, as we have seen, $\langle T_{vv}^{\bar{M}}(\Sigma) \rangle$ starts to become large. The results we obtain there are robust and can be rederived for Schwarzschild spacetime using the results of the Sec. 2.2.

The large fluctuations in the energy density of the infalling matter associated with S-surfaces can drastically affect the background metric as seen by an outside observer. There appears to that observer to be a significant interaction between incoming matter and outgoing Hawking radiation, and this interaction may make the evolution of matter fields appear unitary. On the other hand, for generic foliations in the vicinity of the horizon there is no breakdown in the semiclassical approximation, and no large backreaction, suggesting that an infalling observer falls through the horizon according to semiclassical expectations. The picture that emerges from these very different predictions about the physics of the event horizon is remarkably similar to the ideas underlying the principle of black hole complementarity [6, 7, 11, 12]. The notion of complementarity appears to arise naturally from considering the gravity sector to be quantized and in that sense is not tied to any specific formulation of quantum gravity.

From the perspective of an outside observer, the large fluctuations in energy close to the horizon can be used to delineate the boundary between the region where semiclassical physics is valid, and the region where quantum gravity effects have a dramatic (although largely uncomputable) effect. The location of this boundary, and its appearance to outside observers, is the subject of the next section.

6 Stretched Horizon

It is instructive to evaluate the total energy-momentum tensor in the sense of (29) for large λv :

$$\langle T_{vv}(\Sigma) \rangle = -\frac{\lambda^2}{48\pi} \left[\frac{1}{16(M/\lambda)\alpha^2 e^{2\lambda v} - 1} \right] \quad (64)$$

If on an S-surface we go above a region obeying

$$\lambda x^+ = e^{\lambda v} = \frac{1}{2\alpha\sqrt{M/\lambda}} \quad (65)$$

the induced energy-momentum tensor is small. This region also coincides with the region where the β coefficients are almost zero. It follows that on each S-surface labeled by α there is an approximate value of λv below which there is a problem with the semiclassical approximation (and the concept of a background spacetime is probably absent), but above which there are no local quantum gravity effects. These points, one for each S-surface, define a *timelike* hypersurface marking the boundary between a semiclassical region and a region of strong quantum gravitational effects. Since the approximate value of $\lambda x^- - M/\lambda$ is $-2\alpha\sqrt{M/\lambda}$ for an S-surface defined by α in the vicinity of the horizon, the boundary surface obeys the equation

$$-\lambda x^+ (\lambda x^- - M/\lambda) \approx 1 \quad (66)$$

If we interpret $e^{2\phi}$ to be related to the area of a spherical hypersurface in a four dimensional space time, then by (3) this boundary surface is a constant area hypersurface of one planck unit of area greater than the event horizon (see Fig. 4).

We can estimate the location of the boundary surface between semiclassical and quantum gravitational regions for the Schwarzschild black hole, as seen by an outside observer, using the expression (22) for coordinate shift derived in Sec. 2.2. From equation (21) we see that there is little change in the state for points on an S-surface (now labeled by δ), obeying $\kappa x^+ > 4M\Delta M/\delta$ (remember that the analogue of δ in the CGHS is $\alpha\sqrt{M/\lambda}$). As the value of κx^- for the S-surface throughout this region is approximately δ , we find that the boundary surface is a timelike hypersurface obeying

$$-\kappa x^+ \kappa x^- \approx 4M e \Delta M_{max} \approx 2eM. \quad (67)$$

It follows from equation (13) that this is a constant r hypersurface approximately one Planck length away from the horizon. The distance between the event horizon and this ‘stretched horizon’, in the language of Ref. [20], is determined by the magnitude of the fluctuations of the variables in the gravitational sector (in this case the black hole mass).

For a computation that involves quantities outside the boundary surface, for instance some S matrix elements between states at \mathcal{I}^- and states at \mathcal{I}^+ we can attempt to use the semiclassical approximation, as long as we are not interested in the regions which are to the future of the boundary surface. However, almost all of the Hawking radiation is detected to the future of the boundary surface.

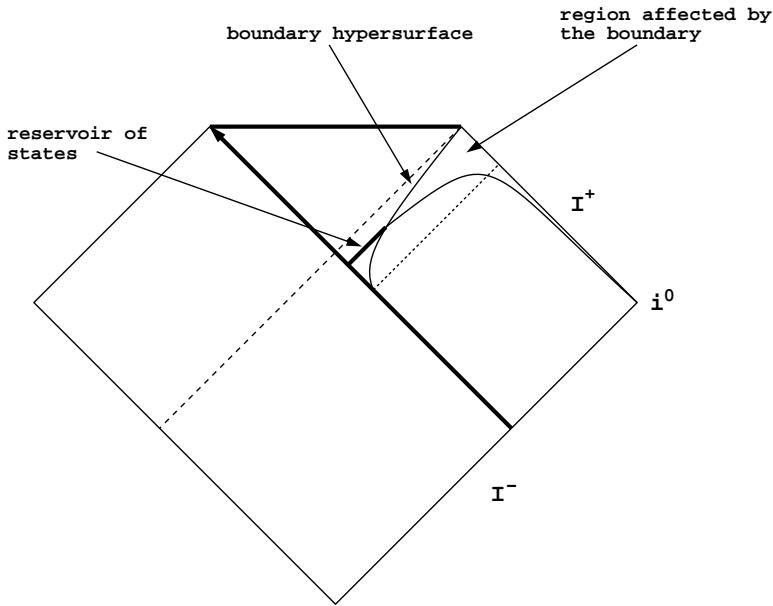


Figure 4: The boundary curve or stretched horizon is shown for the CGHS spacetime.

In order to make any predictions to the future of the boundary surface, we need to use quantum gravity. At this point, there are a couple of possibilities. On the one hand, it may be that the full details of the quantum gravitational interactions behind the stretched horizon are necessary for any calculations to its future, and that no conclusions can be reached about black hole evaporation without a theory that models these interactions. On the other hand, it may be, as has been speculated recently [7], that a theory defined *on* the stretched horizon is sufficient to model all observations made outside the horizon.

Whichever of these scenarios is correct, it is interesting to consider as simple an approach as possible to modeling the physics of the stretched horizon. We should try to work on a single background spacetime, but then we must trace over the microstates of the gravitational field. The in-falling matter states should be treated as being in a statistical ensemble (*i.e.* ρ), so that any outgoing state will interact with this reservoir of states in a very complicated way. As we do whenever a system with a large numbers of degrees of freedom is coupled to a smaller system, we can attempt invoke an effective description of the microstates, using statistical mechanics or thermodynamics. Usually a thermodynamical description is valid if the interaction between the microstates and the small system is weak and if the number of microstates is large enough. In the case of the black hole the second criterion certainly holds as from our calculation one can see that the number of different states at the horizon (the microstates, now thought of as residing in the in-falling matter beyond the boundary surface) is huge. One can then hope to be able to apply statistical mechanics and describe the effective interaction between the gravity fluctuations (that manifest themselves in the form of

a large number of horizon states) and matter fields propagating on the black hole background by ascribing energy, entropy and/or temperature to the boundary surface. It is plausible that this may be the origin of the first law of black hole thermodynamics, and is similar to the stretched horizon idea advocated in [7]. In [16] we shall argue that this procedure associates a temperature to the stretched horizon, and that the log of number of different horizon states should be associated with the Bekenstein entropy of the black hole.

7 Conclusions

We have derived a series of results related to the computation of the state of matter on certain hypersurfaces in a black hole spacetime. These hypersurfaces are characterized by the fact that they yield simultaneous information about the state of matter at future null infinity and close to the event horizon, and are called S-surfaces.

We analyzed the breakdown of the semiclassical approximation for the CGHS model (and for the S-wave matter sector on a Schwarzschild background) on S-surfaces. It was shown in [10] that small fluctuations in the mass of a black hole lead to an entanglement between gravitational and matter degrees of freedom. We have shown here that tracing over the gravitational fluctuations leads to an effective semiclassical description in which a matter state is described by a density matrix rather than a pure state. Each element in the density matrix can be interpreted as the result of a transformation of the original matter state by an operator \mathcal{U} . This operator was constructed explicitly and is closely related to the operator \mathcal{U} defined in Ref. [11] to represent the effect of Planckian interactions near a black hole horizon. We have also shown that there is a large induced energy-momentum tensor associated with the transformed states. This result confirms that there is a dramatic breakdown in the semiclassical picture on S-surfaces, and shows that quantum gravity effects can lead to a large back-reaction close to the horizon. Finally, we identified the region where quantum gravity effects are important, and argued that the boundary of that region is naturally identified as a stretched horizon. Some work related to the results given in this paper can be found in Ref. [21].

The results described are valid only for S-surfaces. A generic foliation of the region close to the horizon would not give rise to a breakdown in the semiclassical approximation. In this sense, the state of matter is not covariant in the usual way, and the complementarity idea arises naturally (see [17] for further discussion of this point). The stretched horizon only has physical meaning for observers that do not cross the horizon.

The quantum gravitational interactions to which we have referred involve degrees of freedom associated with the state of the gravitational background. It is important to note that the quantum nature of the background is closely tied to that of the matter state that originally formed the black hole, since the latter is largely responsible for what we have termed gravitational fluctuations. Thus we can think of there being an entanglement between the degrees of freedom of the matter that created the black

hole and the quantum fields that propagate on the hole. The entanglement gives rise to a reservoir of states that can significantly affect the Hawking radiation. In this paper we have only considered the degree of freedom associated with the energy of the matter that forms the black hole (contained in our discussion in the function $\omega(M)$ of (24)). More generally one should consider the effect of fluctuations in all the degrees of freedom of the matter forming the black hole on in- and out-going quantum fields.

Finally, it is suggested that an outside observer might describe the complex interactions behind the stretched horizon by an effective theory in which matter interacts with a stretched horizon endowed with thermodynamical variables. An effective picture of this kind ignores the details of the microstates behind the stretched horizon so could not lead directly to a unitary picture of black hole evaporation³, but may give a microstate explanation for the laws of black hole thermodynamics.

Acknowledgements

We thank Larry Ford, Esko Keski-Vakkuri and Samir Mathur for helpful discussions.

References

- [1] S. W. Hawking, *Comm. Math. Phys.* **43** (1975) 199; S. W. Hawking, *Phys. Rev. D* **14** (1976) 2460.
- [2] J. Preskill, *Do black holes destroy information?* Presented at the International Symposium on Black holes, Membranes, Wormholes and Superstrings, Woodlands, TX, 16-18 Jan 1992, (hep-th/9209058).
- [3] D. N. Page, *Phys. Rev. Lett.* **44** (1980) 301.
- [4] N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space Times*, Cambridge University Press, Cambridge, UK (1982),
- [5] Y. Aharonov, A. Casher and S. Nussinov, *Phys. Lett.* **191B** (1987) 51.
- [6] G. 't Hooft, *Nucl. Phys.* **B256** (1985) 727; *Nucl. Phys.* **B335** (1990) 138.
- [7] L. Susskind, L. Thorlacius and J. Uglum, *Phys. Rev. D* **48** (1993) 3743; L. Susskind, *Phys. Rev. D* **49** (1994) 6606.
- [8] K. Schoutens, H. Verlinde and E. Verlinde, *Phys. Rev. D* **48** (1993) 2670.
- [9] S. D. Mathur, *Black hole entropy and the semiclassical approximation*, MIT report No. CTP-2304, to appear in the proceedings of the International Colloquium

³One might instead consider assigning the degrees of freedom of the entire reservoir of states to the stretched horizon, perhaps leading to a unitary theory.

on Modern Quantum Field Theory II at TIFR (Bombay), January 1994 (hep-th/9404135).

- [10] E. Keski-Vakkuri, G. Lifschytz, S. D. Mathur and M. E. Ortiz, *Phys. Rev.* **D51** (1995) 1764.
- [11] Y. Kiem, H. Verlinde and E. Verlinde *Quantum Horizons and complementarity*, CERN report CERN-TH-7469-94 (hep-th/9502074).
- [12] C. R. Stephens, G. 't Hooft and B. F. Whiting, *Class. Qu. Grav.* **11** (1994) 621.
- [13] T. Jacobson, *Phys. Rev.* **D48** (1993) 728; F. Englert, S. Massar and R. Parentani, *Class. Quant. Grav.* **11** (1994) 2919; F. Englert, *Operator weak values and black hole complementarity* (gr-qc/9502039).
- [14] C. G. Callan, S. B. Giddings, J. A. Harvey and A. Strominger, *Phys. Rev.* **D45** (1992) R1005.
- [15] J. A. Harvey and A. Strominger, *Quantum aspects of black holes*, in the proceedings of the 1992 TASI Summer School in Boulder Colorado (World Scientific, 1993)
- [16] G. Lifschytz and M. E. Ortiz, *Black hole thermodynamics from quantum gravity*, submitted to Physical Review **D**.
- [17] G. Lifschytz, S. D. Mathur and M. E. Ortiz, *A note on the semiclassical approximation in quantum gravity*, to appear in *Phys. Rev.* **D** (gr-qc/9412040).
- [18] V. Lapchinsky and V. Rubakov, *Acta Phys. Pol.* **B10** (1979) 1041; T. Banks, *Nucl. Phys.* **B249** (1985) 332.
- [19] S. B. Giddings and W. M. Nelson, *Phys. Rev.* **D46** (1992) 2486.
- [20] K.S. Thorne, R.H. Price and D.A. Macdonald, *Black Holes: The Membrane Paradigm*, Yale University Press, New Haven, CT (1986).
- [21] E. Keski-Vakkuri and S. Mathur, in preparation.